SUPPLEMENTAL FILE FOR THE PAPER TITLED "JUMP DETECTION IN GENERALIZED ERROR-IN-VARIABLES REGRESSION WITH AN APPLICATION TO AUSTRALIAN HEALTH TAX POLICIES"

By Yicheng Kang*, Xiaodong Gong†, Jiti Gao‡, and Peihua Qiu*

*University of Florida**,
University of Canberra, Australian National University and IZA†
Monash University‡

In this supplemental file, we mainly give the proof of Theorem 1. First, note that result (i) can be proved similarly to those in cases when there is no measurement error involved (e.g., Qiu *et. al* 1991). Next, let us prove result (ii). Without loss of generality, let us assume that $C = 1$. Consider a given point $x \in [\gamma_1, \gamma_2]$, where $[\gamma_1, \gamma_2]$ is sub-interval in $[0, 1]$. Let $M(x) = m(x)f_X(x)$,

$$\begin{aligned} \widehat{M}_{n,r}(x+) &= \frac{1}{nh_n}\sum_{i=1}^n Y_i K_r\left(\frac{W_i-(x+h_n)}{h_n}\right), \\ \text{and} \qquad \widehat{M}^*_n(x+) &= \frac{1}{nh_n}\sum_{i=1}^n Y_i K_r\left(\frac{W_i-x}{h_n}\right). \end{aligned}$$

Then, it can be verified that, fixing $W_i \in (x, x+h_n)$,

$$\begin{aligned} E\left\{\widehat{M}^*_n(W_i+)|W_i\right\} &= M(x+)\int_0^1 K_r(v)P\left(U < v + \frac{W_i-x}{h_n}\right)\,dv \\ &+ M(x-)\int_0^1 K_r(v)P\left(U \geq v + \frac{W_i-x}{h_n}\right)dv \\ &+ O(h_n+\sigma_n), \\ \text{and} \quad E\widehat{M}_{n,r}(x+) &= M(x+)\int_0^1 K_r(v)P(U < v+1)\,dv \\ &+ M(x-)\int_0^1 K_r(v)P(U \geq v+1)dv \\ &+ O(h_n+\sigma_n). \end{aligned}$$

Next, using some similar arguments to those in the proof of Theorem 3 in

Cheng and Lin (1981), we will show that

$$\widehat{M}_{n,r}(x+) - E\widehat{M}_{n,r}(x+) \quad = \quad O\left(\frac{(\log n)^{1+\eta}}{n^{1/2}h_n}\right), \quad a.s. \tag{A.1}$$

Define the truncated random variable

$$\widetilde{Y}_j = Y_i I_{\{|Y_i|<i^{1/2}\log^{1/2+\gamma}(i)\}}, \qquad i = 1, \dots, n,$$

where $\gamma = (\eta - 1/2)/10$. Let

$$\begin{aligned} g_n(i) &= \frac{1}{nh_n} Y_i K_r\left(\frac{W_i - (x+h_n)}{h_n}\right), \\ \text{and} \qquad \widetilde{g}_n(i) &= \frac{1}{nh_n} \widetilde{Y}_i K_r\left(\frac{W_i - (x+h_n)}{h_n}\right). \end{aligned}$$

Then we have

$$\begin{aligned} & \widehat{M}_{n,r}(x+) - E\widehat{M}_{n,r}(x+) \\ = & \sum_{i=1}^n g_n(i) - \widetilde{g}_n(i) + \sum_{i=1}^n \widetilde{g}_n(i) - E\{\widetilde{g}_n(i)\} + \\ & \sum_{i=1}^n E\{\widetilde{g}_n(i)\} - E\{g_n(i)\}. \end{aligned}$$

Consider, for any given $\varepsilon > 0$,

$$\begin{aligned} & P\left(\frac{\xi_n}{(\log n)^{1+\gamma}} \sum_{i=1}^n [\widetilde{g}_n(i) - E\widetilde{g}_n(i)] > \varepsilon\right) \\ = \; & P\left(\xi_n \sum_{i=1}^n [\widetilde{g}_n(i) - E\widetilde{g}_n(i)] > \varepsilon(\log n)^{1+\gamma}\right) \\ \leq \; & \exp\left\{-\varepsilon(\log n)^{1+\gamma}\right\} E\left\{\exp\left\{\xi_n \sum_{i=1}^n [\widetilde{g}_n(i) - E\widetilde{g}_n(i)]\right\}\right\} \\ = \; & n^{-\varepsilon(\log n)^\gamma} \prod_{i=1}^n E\left\{\exp\left\{\xi_n[\widetilde{g}_n(i) - E\widetilde{g}_n(i)]\right\}\right\}, \qquad \text{(A.2)} \end{aligned}$$

where $\xi_n = n^{1/2}h_n(\log n)^{-1/2-2\gamma}$. By the definition of $\widetilde{g}_n(i)$ and the boundedness of K_r, we have

$$\begin{aligned} \text{(A.3)} \qquad & |\xi_n[\widetilde{g}_n(i) - E\widetilde{g}_n(i)]| \\ = \; & \frac{\xi_n}{nh_n}\left|\widetilde{Y}_i K_r\left(\frac{W_i - (x+h_n)}{h_n}\right) - E\left[\widetilde{Y}_j K_r\left(\frac{W_i - (x+h_n)}{h_n}\right)\right]\right| \\ = \; & \frac{\xi_n}{nh_n} n^{1/2}\log^{1/2+\gamma}(n) C_1 \\ = \; & (\log n)^{-\gamma} C_1, \end{aligned}$$

where C_1 is some positive constant. Using the facts that $\exp(x) \le 1+x+x^2$, for $|x| \le 1/2$, and (A.3), we have

$$\begin{aligned}
\text{(A.2)} \quad &\le n^{-\varepsilon(\log n)^\gamma} \prod_{i=1}^n E\left\{1+\xi_n^2[\widetilde{g}_n(i) - E\widetilde{g}_n(i)]^2\right\} \\
&= n^{-\varepsilon(\log n)^\gamma} \prod_{i=1}^n \left\{1+\xi_n^2 Var[\widetilde{g}_n(i)]\right\} \\
&\le n^{-\varepsilon(\log n)^\gamma} \prod_{i=1}^n \exp\left\{\xi_n^2 Var[\widetilde{g}_n(i)]\right\} \\
&= n^{-\varepsilon(\log n)^\gamma} \exp\left\{\sum_{i=1}^n \xi_n^2 Var[\widetilde{g}_n(i)]\right\} \\
&\le n^{-\varepsilon(\log n)^\gamma} \exp\left\{\frac{\xi_n^2}{n^2 h_n^2} nE\left[\widetilde{Y}_1^2 K_r^2\left(\frac{W_1-(x+h_n)}{h_n}\right)\right]\right\} \\
&\le n^{-\varepsilon(\log n)^\gamma} \exp\left(\frac{\xi_n^2}{nh_n^2} C_2'\right) \\
&\le C_2 n^{-\varepsilon(\log n)^\gamma}, \qquad \text{(A.4)}
\end{aligned}$$

where C_2' and C_2 are two positive constants. In the second last inequality, we have used the boundedness of $K_r(\cdot)$, $a(\cdot)$, $b'(\cdot)$ and $b''(\cdot)$, and the fact that $E[\widetilde{Y}_1^2] \le E[Y_1^2] = b'(\theta)^2 + b''(\theta)a(\phi)$. By the same arguments in (A.2)-(A.4), we have

$$P\left(\frac{\xi_n}{(\log n)^{1+\gamma}} \sum_{i=1}^n [E\widetilde{g}_n(i) - \widetilde{g}_n(i)] > \varepsilon\right) \le C_2 n^{-\varepsilon(\log n)^\gamma}.$$

And thus,

$$P\left(\left|\frac{\xi_n}{(\log n)^{1+\gamma}} \sum_{i=1}^n [E\widetilde{g}_n(i) - \widetilde{g}_n(i)] > \varepsilon\right|\right) \le 2C_2 n^{-\varepsilon(\log n)^\gamma}.$$

By the Borel-Cantelli Lemma, we have

$$\frac{\xi_n}{(\log n)^{1+\gamma}} \sum_{i=1}^n [\widetilde{g}_n(i) - E\widetilde{g}_n(i)] \to 0, \qquad a.s. \tag{A.5}$$

Next, it can be checked that

$$\left|\frac{\xi_n}{(\log n)^{1+\gamma}} \sum_{i=1}^n [E\widetilde{g}_n(i) - Eg_n(i)]\right| \tag{A.6}$$

$$
\begin{aligned}
&\le \frac{C_3'\xi_n}{nh_n(\log n)^{1+\gamma}} \sum_{i=2}^{n} E|Y_i| I_{\{|Y_i| \ge i^{1/2}(\log i)^{1/2+\eta}\}} \\
&\le \frac{C_3'\xi_n}{nh_n(\log n)^{1+\gamma}} \sum_{i=2}^{n} EY_i^2/(i^{1/2}(\log i)^{1/2+\eta}) \\
&\le \frac{C_3\xi_n}{nh_n(\log n)^{1+\gamma}} n^{1/2} \\
&= \frac{C_3\xi_n}{n^{1/2}h_n(\log n)^{1+\gamma}} \\
&= C_3(\log n)^{-3/2-3\gamma},
\end{aligned}
$$

where C_3' and C_3 are two positive constants. Also,

$$
\left| \frac{\xi_n}{(\log n)^{1+\gamma}} \sum_{i=1}^{n} [g_n(i) - \widetilde{g}_n(i)] \right| \le \frac{C_4\xi_n}{nh_n(\log n)^{1+\gamma}} \sum_{i=1}^{n} |Y_i - \widetilde{Y}_i|.
$$

Notice that

$$
\begin{aligned}
E\sum_{k=1}^{\infty} I_{\{Y_k \ne \widetilde{Y}_k\}} &= \sum_{k=2}^{\infty} P(|Y_k| \ge k^{1/2}(\log k)^{1/2+\gamma}) \\
&\le 1 + \sum_{k=2}^{\infty} EY_k^2/(k(\log k)^{1+2\gamma}) \\
&< \infty.
\end{aligned}
$$

So, for a given ω, there exists an $N(\omega) > 0$ such that $Y_k(\omega) = \widetilde{Y}_k(\omega)$, for $k \ge N(\omega)$. Then we have

$$
\begin{aligned}
\left| \frac{\xi_n}{(\log n)^{1+\gamma}} \sum_{i=1}^{n} [g_n(i) - \widetilde{g}_n(i)] \right| &\le \frac{\xi_n C_4(\omega)}{nh_n(\log n)^{1+\gamma}} \\
&= \frac{C_4(\omega)}{n^{1/2}(\log n)^{3/2+3\gamma}},
\end{aligned}
\tag{A.7}
$$

where C_4 is a positive constant. By combining (A.5) - (A.7), we have shown that (A.1) is true. By arguments similar to (A.5) - (A.7), we also have

$$
\widehat{M}_n^*(W_i+) - E[\widehat{M}_n^*(W_i+)|W_i] = O\left(\frac{(\log n)^{1+\eta}}{n^{1/2}h_n} \right), \quad a.s.
$$

Next, consider the special case when $Y_i \equiv 1$, for $i = 1, \cdots, n$. In this case, by replacing Y_i with 1 in the definition of $g_n(i)$, we have

$$
\begin{aligned}
g_n(i) &= \frac{1}{nh_n} K_r\left(\frac{W_i - (x + h_n)}{h_n} \right), \\
\text{and} \qquad \widetilde{g}_n(i) &= g_n(i), \text{ for } i = 1, \ldots, n.
\end{aligned}
$$

Then, repeating the arguments in (A.2) - (A.5) gives

$$\begin{aligned} & \frac{1}{nh_n}\sum_{j=1}^{n} K_r\left(\frac{W_j-(x+h_n)}{h_n}\right) - E\left[\frac{1}{nh_n}\sum_{j=1}^{n} K_r\left(\frac{W_j-(x+h_n)}{h_n}\right)\right] \\ = \; & O\left(\frac{(\log n)^{1+\eta}}{n^{1/2}h_n}\right), \quad a.s. \end{aligned}$$

By the continuity of f_X, we have,

$$\frac{1}{nh_n}\sum_{j=1}^{n} K_r\left(\frac{W_j-(x+h_n)}{h_n}\right) = f_X(x) + O\left(\beta_n\right), \quad a.s.$$

and

$$\frac{1}{nh_n}\sum_{j=1}^{n} K_r\left(\frac{W_j-W_i}{h_n}\right) = f_X(x) + O\left(\beta_n\right), \quad a.s.$$

Since f_X is positive on $(0,1)$, we have

$$\begin{aligned} & \max_{x<W_i<x+h_n} |\widehat{m}_n^*(W_i+) - \widehat{m}_{n,r}(x+)| \\ = \; & |m(x+)-m(x-)| \int_0^1 K_r(v) P\left(v + \min_{x<W_i<x+h_n} \frac{W_i-x}{h_n} < U < v+1\right) dv \\ & +O(\beta_n), \; a.s. \\ = \; & |m(x+)-m(x-)| \int_0^1 K_r(v) P(v<U<v+1)\, dv + O(\beta_n), \quad a.s. \end{aligned}$$

Next, let

$$\begin{aligned} & \widehat{M}_n(x+) \\ = \; & \frac{1}{nh_n}\sum_{i=1}^{n} Y_i K_r^*\left(\frac{W_i-x}{h_n}\right) K^*\left(\frac{|\widehat{m}_n^*(W_i+) - \widehat{m}_{n,r}(x+)|}{\max_{x<W_i<x+h_n} |\widehat{m}_n^*(W_i+) - \widehat{m}_{n,r}(x+)|}\right). \end{aligned}$$

Then, we have

$$\widehat{M}_n(x+) = \frac{1}{nh_n}\sum_{i=1}^{n} Y_i K_r^{**}\left(\frac{W_i-x}{h_n}\right) + O(\beta_n), \quad a.s.$$

We can check that

$$
\begin{aligned}
& E\left[\frac{1}{nh_n}\sum_{i=1}^{n} Y_i K_r^{**}\left(\frac{W_i - x}{h_n}\right)\right] \\
= & \int\int M(wh_n + x - \sigma_n u) K_r^{**}(w) f_U(u)\, du dw \\
= & \int_0^1 K_r^{**}(w)\, dw \left\{ \int_{-\infty}^{\frac{wh_n}{\sigma_n}} [M(x+) + M'(x + \xi_1(wh_n - \sigma_n u)) \right. \\
& (wh_n - \sigma_n u)] f_U(u)\, du + \int_{\frac{wh_n}{\sigma_n}}^{\infty} [M(x-) + M'(x + \xi_2(wh_n - \sigma_n u)) \\
& \left. (wh_n - \sigma_n u)] f_U(u)\, du \right\} \\
= & M(x+) \int_0^1 K_r^{**}(w) P(U < wh_n/\sigma_n)\, dw \\
& + M(x-) \int_0^1 K_r^{**}(w) P(U > wh_n/\sigma_n)\, dw + O(h_n) + O(\sigma_n) \\
= & M(x+) \int_0^1 K_r^{**}(w)\, dw - [M(x+) - M(x-)] \int_0^1 K_r^{**}(w) P(U > w)\, dw \\
& + O(h_n),
\end{aligned}
$$

where ξ_1 and ξ_2 are two values in $(0, 1)$. By the arguments analogous to those in (A.2) - (A.7), it can be proved that

$$
\begin{aligned}
\widehat{M}_n(x+) = & M(x+) \int_0^1 K_r^{**}(w)\, dw - [M(x+) - M(x-)] \cdot \\
& \int_0^1 K_r^{**}(w) P(U > w)\, dw + O(\beta_n), a.s.
\end{aligned}
$$

After applying the above arguments to the special case when $Y_i \equiv 1$, for $i = 1, \cdots, n$, we have

$$
\begin{aligned}
& \frac{1}{nh_n}\sum_{i=1}^{n} K_r^{*}\left(\frac{W_i - x}{h_n}\right) K^{*}\left(\frac{|\widehat{m}_n^{*}(W_i+) - \widehat{m}_{n,r}(x+)|}{\max_{x<W_i<x+h_n} |\widehat{m}_n^{*}(W_i+) - \widehat{m}_{n,r}(x+)|}\right) \\
= & f_X(x) \int_0^1 K_r^{**}(w)\, dw + O(\beta_n), \quad a.s.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\widehat{m}_n(x+) - \widehat{m}_n(x-) = & m(x+) - m(x-) - (m(x+) - m(x-)) \cdot \\
& \frac{\int_0^1 K_r^{**}(w) P(|U| > w)\, dw}{\int_0^1 K_r^{**}(w)\, dw} + O(\beta_n), \quad a.s.
\end{aligned}
$$

Next, observe that $K^*\left(\frac{\int_0^1 K_r(v)P(v+w<U<v+1)\ dv}{\int_0^1 K_r(v)P(v<U<v+1)\ dv}\right)$ increases with w, $P(|U| > w)$ decreases with w, and K_r^* is a decreasing function. Then, the inequality given in the last paragraph of Section 3 follows from the Chebyshev integral inequality[1]. So, we have completed the proof of result (ii). To prove result (iii), note that the convergences in (i) and (ii) are uniform in $x \in [\gamma_1, \gamma_2]$ and that

$$\frac{\int_0^1 K_r^{**}(w)P(|U| > w)\ dw}{\int_0^1 K_r^{**}(w)\ dw} < 1$$

by the condition that $\int_{-\delta}^{\delta} f_U(u)\ du > 0$ for any $\delta > 0$. Then, the result (iii) follows immediately. The proof of Theorem 1 is then completed.

[1]Let $f, g : [a, b] \to \mathcal{R}$ be integrable functions, both increasing or both decreasing. Furthermore, let $p : [a, b] \to \mathcal{R}_0^+$ be an integrable function. Then

$$\int_a^b p(x)\ dx \int_a^b p(x)f(x)g(x)\ dx \geq \int_a^b p(x)f(x)\ dx \int_a^b p(x)g(x)\ dx.$$

See Mitrinovic *et. al* (1993) for a detailed discussion on Chebyshev Integral Inequality and its generalizations and variants.